FRONT MATTER

Title

Optimized gene expression from bacterial chromosome by high-throughput integration and screening

Short title: Optimizing expression by Tn5 integration

Authors

Tatyana E. Saleski[1], Meng Ting Chung[2], David N. Carruthers[1], Azzaya Khasbaatar[1], Katsuo Kurabayashi[2,3], Xiaoxia Nina Lin[1,4]*

Affiliations

[1]Department of Chemical Engineering, University of Michigan, Ann Arbor, MI, USA.
[2]Department of Mechanical Engineering, University of Michigan, Ann Arbor, MI, USA.
[3]Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA.
[4]Department of Biomedical Engineering, University of Michigan, Ann Arbor, MI, USA.

*Corresponding author. Email: ninalin@umich.edu.

Abstract

Chromosomal integration of recombinant genes is desirable compared to expression from plasmids due to increased stability, reduced cell-to-cell variability, and the elimination of antibiotics for plasmid maintenance. Here, we present a new approach for tuning pathway gene expression levels via random integrations and high-throughput screening. We demonstrate multiplexed gene integration and expression-level optimization for isobutanol production in *Escherichia coli*. The integrated strains could, with significantly lower expression levels than plasmid-based expression, produce high titers (10.0 ± 0.9 g/L isobutanol in 48 h) and yields (69 % of the theoretical maximum). Close examination of pathway expression in the top-performing, as well as other isolates, reveals the complexity of cellular metabolism and regulation, underscoring the need for precise optimization while integrating pathway genes into the chromosome. We expect this new method for multiplexed pathway gene integration and expression optimization can be readily extended to a wide range of pathways and chassis to create robust and efficient production strains.

38 **MAIN TEXT**
39
40 **The manuscript should be a maximum of 15,000 words.**
41
42 **Introduction**
43
44 Microbial biosynthesis is a sustainable, high-specificity approach to achieving chemical
45 conversions with the potential to produce a vast assortment of pharmaceutical, fuel, and commodity
46 chemicals. Development of a high-performing production strain for a desired molecule generally
47 requires tuning the expression levels of native and/or heterologous genes in hosts such as
48 *Escherichia coli* (*1, 2*). While extrachromosomal multicopy plasmids offer a convenient method for
49 rapidly prototyping different expression levels of genes, and the majority of metabolic engineering
50 efforts utilize them, they have several challenging attributes. First and foremost, plasmids are
51 unstable, which has been extensively documented and studied due to the importance of plasmids in
52 biotechnology (*3-9*). Cells with plasmids can suffer from both structural instability, in which the
53 plasmid is still carried but mutations inactivate the gene of interest, as well as segregational
54 instability, in which some cells no longer carry the plasmid (*10, 11*). Plasmids can also multimerize,
55 which decreases their stability (*12, 13*). Second, plasmid copy number can vary between cells within
56 stable populations, and the degree of variation is often not well characterized even for widely used
57 plasmids (*14*). Münch *et al.* (*15*) presented a striking example of cellular heterogeneity during
58 plasmid-based production of recombinant protein in *Bacillus megaterium*. The authors found that,
59 under the condition of strong selective pressure, 30% of the cell population was in a low-producing
60 state. This heterogeneity was found to be a result of asymmetric plasmid distribution. Moreover,
61 plasmids require a selective pressure, which is typically an antibiotic molecule. The addition of
62 antibiotics raises the process costs, puts additional stress on the cells, and is not always effective in
63 enforcing maintenance of the plasmid. Kanamycin, for example, can become an ineffective selection
64 agent at high phosphate concentrations (*16*). Furthermore, due to the rise of multidrug-resistant
65 pathogens, there is strong motivation to reduce antibiotic usage (*17*).

66
67 Chromosomal integration and expression of genes is an effective way of avoiding the issues
68 of plasmids described above. This alternative approach leads to more stable and robust production
69 strains, as well as reduce leakiness of inducible expression systems, which are highly desirable traits
70 for large-scale and long-term production processes (*18, 19*). Mairhofer *et al.* (*20*) compared a
71 plasmid-based to a chromosomally integrated expression system and found a much more significant
72 stress response in the plasmid-carrying strain. They identified massive overtranscription of the
73 plasmid-based gene of interest, leading to diversion of ribosomes and other cellular resources, as
74 the source of the metabolic burden.

75 Significant research efforts have been devoted to developing techniques for synthetic
76 integrations (*21-23*). The two tools most widely used in prokaryotes for targeted integration of
77 constructs are homologous recombination by expression of the RecET or λ-Red proteins (*24-26*)
78 and site-specific recombination (*27*). Homologous recombination enables rapid integration into a
79 desired locus but is only suitable for constructs up to several kilobases in size and that do not carry
80 significant homology to another part of the genome. Site-specific recombination is less restrictive
81 of the input construct but requires a bacterial chromosomal attachment site (*attB*) in the recipient
82 genome, which can place limitations on the location of integration. This site restrictiveness can be
83 circumvented by first inserting the attachment site into the target locus (*28*). Yet, despite the
84 recognized desirability of chromosomal integration and the genetic tools available for targeted
85 integrations, achieving suitable gene expression levels from the chromosome for optimal production
86 remains far more challenging than manipulation of gene expression via plasmids, even in well-
87 studied organisms such as *E. coli* (*22*).

88 Chromosomal gene expression levels are dependent on multiple determinants. Similar to
89 plasmid-based gene expression, elements of the genetic construct, including promoters, ribosome-
90 binding sites (RBSs), and enhancers or activators, can be exploited to optimize heterologous gene
91 expression. Yet, single-copy expression from the chromosome is generally weaker than that of the

92 same construct from a multi-copy plasmid. Subsequently, there have been efforts in increasing the
93 copy number on the chromosome to achieve high-level gene expression (*18, 29*).

94 Another determinant affecting the expression level from the chromosome strongly is the
95 location of gene integration. It has long been known that during exponential growth, genes closer to
96 the origin of replication experience higher gene dosage and, therefore, higher expression levels (*30,*
97 *31*). More recently it has been shown that other factors related to genomic position, such as level of
98 DNA compaction or proximity to active genes, can also affect gene expression levels (*32, 33*). In
99 examining transcription levels of a reporter gene at various sites in the *E. coli* genome, Bryant *et al.*
100 (*32*) and Scholz *et al.* (*33*) observed significant differences, up to ~300-fold, in expression across
101 the *E. coli* genome, excluding gene dosage effects. Intriguingly, on the other hand, it has been
102 suggested that in nature, microorganisms may use transposon-mediated transfer of catabolic genes
103 to different genomic locations as a means of fine-tuning their expression levels in order to adapt to
104 new environmental conditions (*34*). These findings suggest that modulation of gene integration
105 position can serve as a useful tool for regulating expression levels. For instance, Loeschcke *et al.*
106 (*35*) have developed an approach for transfer and expression (TREX) of biosynthetic pathways in
107 bacteria utilizing Tn5 transposase to integrate gene clusters into random sites of the chromosomes
108 and, based on resulting colony color intensities, observed a range of production levels for pigmented
109 secondary metabolites.

110 Synthetic biology applications targeting the production of a small molecule often require
111 delicately balanced cellular resource allocation between production and growth, as well as across
112 different genes of the biosynthetic pathway. It remains elusive to predict the optimal gene expression
113 levels for most pathways due to the highly entangled nature of cellular metabolic and regulatory
114 networks. Additionally, for chromosomal gene expression, it is not fully understood yet how the
115 expression of different regions of the chromosome is affected by changes in environmental
116 conditions and whether/how an inserted gene construct will affect the native expression level of the

117 insertion region. It is, therefore, very difficult to rationally decide where to integrate a gene
118 construct. Here, we present a new metabolic engineering strategy for optimizing a pathway's
119 performance by tuning the gene expression through position-dependent expression variation (Fig.
120 1). In this method, we utilize the Tn5 transposase to randomly integrate pathway genes in the *E. coli*
121 genome in a multiplexed fashion. We subsequently screen the libraries using syntrophic co-culture
122 amplification of production (SnoCAP), which converts the production phenotype into a readily
123 screenable growth phenotype, as described in our previous work (*36*). We demonstrate the
124 approach using a pathway for the production of isobutanol, a promising drop-in biofuel (*37*).

125 Efficient isobutanol production in *E. coli* has been achieved by diverting the branched-chain
126 amino acid pathway intermediate 2-ketoisovalerate (2-KIV) for conversion into isobutanol (*38*)
127 (Fig. 2A). It has been demonstrated that by overexpressing AlsS from *Bacillus subtilis*, IlvC and
128 IlvD from *E. coli* (for the conversion of pyruvate to 2-KIV), and KivD and AdhA from *Lactococcus*
129 *lactis* (for the conversion of 2-KIV to isobutanol), production can reach up to 84% of the theoretical
130 yield under aerobic batch conditions (*38*) and titers can reach up to 50.8 g/L under fed-batch
131 conditions (*39*). Resolution of a cofactor imbalance has resulted in strains reaching 100% of
132 theoretical yield under anaerobic conditions (*40*), and a recently developed strategy for isobutanol-
133 linked growth selection under anaerobic conditions enabled optimization of pathway expression
134 levels for achieving optimal production rates (*41*).

135 While much work with this pathway has relied on plasmid-based expression of the
136 pathway genes, there is also interest in developing chromosomally integrated strains. Akita *et al.*
137 (*42*) have demonstrated the integration of the five pathway genes into the *E. coli* genome, resulting
138 in a titer of 6.8 g/L and a yield of 55%, without the addition of costly antibiotics or inducers.
139 Bassalo *et al.* (*43*) developed a new CRISPR-based method for high-efficiency, single-step
140 chromosomal integration in *E. coli* and, in demonstrating its significant size capability, reported

141 integration of the entire five-gene isobutanol pathway into a single chromosomal locus, leading to
142 2.2 g/L isobutanol from 85 g/L glucose.
143
144 In the present work, we demonstrate that the generation of libraries with the isobutanol
145 pathway genes integrated into diverse chromosomal positions combined with high-throughput
146 production phenotype screening is an effective means of generating high-performance
147 chromosomally integrated strains. *E. coli* strains resulted from this study show, to our knowledge,
148 the highest titers and yields yet reported with the isobutanol pathway genes expressed exclusively
149 from the chromosome.

150 **Results**

151

152 *Construction and screening of Tn5 integration libraries*

153 We began by constructing an integration library (Library A) in which a construct containing
154 the *alsS* gene, under the PLlacO1 promoter, and a kanamycin-resistance gene was integrated by Tn5
155 transposase into the genome of JCL260 Δ*lysA* in random locations. JCL260 was developed by
156 Atsumi *et al.* (*38*) as an isobutanol production strain and its genome features six gene deletions
157 created to decrease byproduct formation, as well as the *lacI*^{q} mutation for reduced leakiness of genes
158 under the *lac* promoter prior to induction. We introduced the Δ*lysA* deletion into JCL260 in our
159 previous work in order to render the strain a lysine auxotroph for applying the SnoCAP screening
160 approach (*36*). In the SnoCAP approach, a library is co-encapsulated in water-in-oil microdroplets
161 with a fluorescent sensor strain that is auxotrophic for the target molecule of interest. The library
162 itself is also auxotrophic for a second, orthogonal molecule (here, lysine) supplied by the sensor
163 strain. This cross-feeding configuration enables conversion of the production phenotype into a
164 fluorescent output, amplified by co-culture growth (*36*). To screen for overproducers of 2-
165 ketoisovalerate, an intermediate in the isobutanol pathway (Fig. 2A), we use a sensor strain with a
166 deletion of the *ilvD* gene (encoding a dihydroxy-acid dehydratase), which prevents it from
167 converting 2,3-dihydroxy-isovalerate into 2-ketoisovalerate. This strain can only grow when it is
168 provided an exogenous source of 2-ketoisovalerate and/or branched-chain amino acids. We
169 demonstrated in our previous work that this sensor can be used to identify strains that will
170 overproduce isobutanol once the genes responsible for converting 2-ketoisovalerate to isobutanol
171 (*kivD* and *adhA*) are added to them (*36*).

172 The aforementioned integration of *alsS* into JCL260 Δ*lysA* was accomplished by PCR
173 amplification of the *alsS* and *kan* genes with primers that add the 19 bp mosaic end sequence
174 recognized by Tn5, followed by reaction with Tn5, and electroporation of cells with the resulting
175 transposome. The approximate library size was determined to be ~$3.4 \pm 1.1 \times 10^4$ by plating a

176 small quantity of the cells after a short initial recovery on kanamycin-containing plates. The
177 remainder of the library was subjected to a liquid kanamycin selection and subsequently screened
178 by SnoCAP.

179 The cells are loaded into microdroplets at densities such that all droplets contain sensor cells,
180 and most droplets contain either zero or one secretor cell from the integration library. Thus, the co-
181 growth phenotype of individual library members may be assessed. The microdroplet
182 compartmentalization format enables high-throughput screening and isolation of droplets of interest
183 by fluorescence-activated droplet sorting (FADS) (*44*).

184 Fluorescence measurement of the droplets containing Library A members co-cultured with
185 fluorescent sensor strain showed that more droplets exhibited growth than when droplets contained
186 co-cultures with the parent strain, which lacks the *alsS* gene, as the secretor (Fig. 2Bi,ii). We sorted
187 the top 0.3% of droplets with the highest fluorescence signal and plated the pool of droplets on plates
188 with tetracycline to select for the secretor cells and against the sensor cells (the parent strain of the
189 secretor strain libraries contains a tetracycline resistance gene in the F plasmid, whereas the sensor
190 strain is tetracycline sensitive). We assessed the isobutanol production of colonies obtained from
191 the screening and colonies from the unscreened library, after transformation with plasmid pSA65,
192 which encodes KivD and AdhA. Of 12 colonies chosen at random from the unscreened library, 10
193 produced less than 0.1 g/L after 48 h of fermentation (similar to the parent strain), one produced 0.4
194 g/L and the other produced 2.1 g/L (Fig. S1). Of 10 randomly selected colonies from the sorted pool,
195 all produced more than 2.5 g/L (Fig. S1). There is a significant difference in isobutanol production
196 between these two groups (P-value < 0.0001, unpaired two-sample t-test), which demonstrates the
197 effectiveness of the screening method.

198 We also attempted integration of the entire three gene operon (*alsS, ilvCD*), along with the
199 kanamycin resistance cassette, in one transposome. This is a 6.3 kb construct, compared to 3 kb for
200 *kan-alsS*. While we could generate a substantial number of kanamycin-resistant cells from the whole

201 operon integration ($6.0 \pm 2.7 \times 10^3$ in a single integration), upon encapsulation with the sensor strain
202 and incubation, we did not detect any co-growth in the droplets at either 0.5 or 1.0 g/L norvaline,
203 whereas the *kan-alsS* integration library led to droplets displaying significant cell growth at both of
204 these norvaline concentrations. We hypothesized this lack of growth could be due to a problem with
205 the library generation, such as incomplete integrations.

206 Due to the above observation that the longer construct had a lower integration efficiency, we
207 chose to minimize construct length and use the *ilvCD* genes as a selection marker by utilizing their
208 essentiality in minimal medium, rather than an antibiotic. We deleted *ilvD* from JCL260 Δ*lysA* and
209 integrated constructs containing either both genes *ilvC* and *ilvD* behind the PLlacO1 promoter or
210 only the *ilvD* gene behind the same promoter. We performed the selection in minimal M9 medium
211 supplemented with IPTG and lysine, with a small amount of the transformed cells plated on solid
212 medium for library size estimation and the rest undergoing liquid selection. These integrations
213 produced far smaller library sizes than the *alsS* integration construct (~10^2 integrants per
214 electroporation vs. ~10^4 for *alsS*). We determined that this was due to substantially lower survival
215 rates in the minimal medium selection compared to kanamycin selection in LB medium. For
216 example, when JCL260 Δ*lysA* Δ*ilvD* was transformed with pSA69, which carries *alsS/ilvCD*, as
217 well as a kanamycin resistance gene, the number of colonies recovered on LB plates with kanamycin
218 was two orders of magnitude higher than the number on M9 plates with lysine and IPTG. Further
219 optimization of the recovery method, such as supplementation of additional nutrients or seeding of
220 small quantities of the branched-chain amino acids may help improve this lower-than-expected
221 recovery rate and increase the diversity of the library. It should also be noted that the survival rate
222 on the solid medium may be different from that in the liquid medium, which could contribute to
223 inaccuracies in estimated library size. Deep sequencing of the library could be used to quantify the
224 library size more accurately. We repeated the *ilvD* and *ilvCD* library generation several times and
225 pooled the integrants, resulting in libraries of each of size ~500-1,000 (estimated from agar plating).

226 We then integrated the *alsS-kan* transposome into these libraries, which led to ~1.1 ± 0.5 x 10^4
227 integrants for the *ilvD/alsS* library (Library B) and 1.3 ± 0.3 x 10^4 for the *ilvCD/alsS* library (Library
228 C). Therefore, the library should consist of about 10 different *alsS* integration locations for each
229 *ilvD* or *ilvCD* integration library member.

230 When assessed by the SnoCAP method (library members co-encapsulated in microdroplets
231 with sensor strain, incubated for co-growth, and the fluorescence of the droplets measured), the *ilvD*
232 and *ilvCD* integration libraries displayed no measurable co-growth. The *ilvD/alsS* and *ilvCD/alsS*
233 libraries (Libraries B and C, respectively) on the other hand, showed substantially more fluorescence
234 compared to the library with only *alsS* integrated (Library A) (Fig. 2B).

235 We sorted the droplets containing co-cultures of these libraries and recovered colonies from
236 the pools of droplets, and then transformed the isolates with pSA65 and assessed isobutanol
237 production. The best isolates we identified came from Library C. In particular, an isolate termed C7
238 reached a production level closely approaching that of the strain expressing the pathway genes from
239 a plasmid (Fig. 3A). With the goal of understanding the relationship between production phenotype
240 and genome-position-dependent gene expression, we decided to examine isolates displaying a
241 spectrum of production levels and hence also isolated several library members with low- and
242 intermediate-production levels (Fig. 3A) by screening the random libraries via the microplate format
243 of SnoCAP.

244

245 *Genomic characterization of select library isolates*
246
247 We determined the insertion site locations in the selected isolates by transposon footprinting,
248 which also allowed us to eliminate duplicates of the same genotype and resulted in a panel of unique
249 isolates reported in this work. The integration locations were found to spread across the whole
250 genome (Fig. 3B, Table 1). Intriguingly, each of the Library B isolates, B1 and B2, contains
251 integrations into the F plasmid, a large (~100 kb) plasmid previously introduced to the base strain

252 to supply *lacIq* for increased levels of the *lac* repressor. The F plasmid is stably maintained in *E.*
253 *coli* due to an active partitioning system (*45, 46*), and it has been suggested that placing
254 heterologous genes on the F plasmid may provide improved stability over the use of other
255 plasmids (*47, 48*). It is also interesting that in isolate B2, both integration constructs are found to
256 be located on the F plasmid.

257 Another noteworthy observation was that two of the disrupted genes, *ttdT* and *yjhI*, are genes
258 in which we had found mutations during previous work in which we chemically mutagenized
259 JCL260 Δ*lysA*, screened for improved isobutanol productivity and resequenced the genome of an
260 improved isolate (*36*). In that study, the strain with the best production level possessed a mutation
261 in the *ttdT* gene leading to an alanine to valine substitution at position 390 of the protein, and
262 *yjhI* contained a mutation leading to a premature stop codon after the first 187 amino acids (of
263 227).

264 In the best isolate, C7, the two integration constructs landed in *ttdT* and *yqiH,* two genes
265 that are quite close to each other, separated by only 18 kb (Fig. 4A). To better understand the genetic
266 mechanism underlying the high-production phenotype of this isolate, we investigated whether the
267 gene disruptions, besides expression of the inserted genes, can contribute to increased production.
268 We knocked out the *ttdT* and *yqiH* genes individually in a base strain JCL260 Δ*lysA*. Into each
269 deletion strain, we then integrated *alsS* in the *yghX* site, which we have previously seen leads to an
270 intermediate production level (and therefore should allow us to see changes in the production that
271 may not be visible in a very low- or high-producing background strain). After transformation of the
272 strains with pSA65, we tested isobutanol production and found that the *ttdT* deletion strain produces
273 more isobutanol to a statistically significant degree compared to the strain possessing wildtype *ttdT*,
274 whereas *yqiH* deletion does not have a statistically significant effect (Fig. 4B).

275
276
277
278

279 *Characterization of pathway gene transcript levels in select library isolates*

280 We next analyzed gene expression at the transcript-level in various isolates of different
281 production levels. We harvested the RNA during exponential growth (3 h after induction of the
282 pathway genes with IPTG), reverse transcribed it, and analyzed gene expression of *alsS*, *ilvC*,
283 and *ilvD* by quantitative real-time PCR (qPCR) (Fig. 5B), while continuing to grow the cultures
284 and measuring their isobutanol production 48 h following induction (Fig. 5A). It was found that
285 the pathway gene expression in each of the integrated strains was, as expected, far lower than
286 that in the strain expressing *alsS/ilvCD* from the pSA69 plasmid (Fig. 5B, note the log-scale of
287 the Y-axis). Yet, with significantly lower expression levels of all three genes, the top isolate C7
288 showed a production level very close to that of the plasmid strain. Although random integration
289 may affect cell growth, all of the examined isolates displayed similar growth patterns under the
290 culture conditions used in this study (Fig. S2). This is likely because integrants with significant
291 growth defects were lost during the stages of the library construction and screening process that
292 involved monoculture growth (liquid selection after transposon integration and selection on agar
293 plates after SnoCAP screening).

294 We noted from the quantified transcript levels that there was a lack of strong correlation
295 between the expression level and the distance from the origin. For instance, close examination of
296 the four isolates from Library A (A1-A4) showed that there was a large variation in the expression
297 level of *alsS* between these isolates, whose integration sites were at 3.8, 4.5, 4.2, and 1.4 Mb,
298 respectively (Fig. 3B). Yet, the expression level of A1 was very low despite the integration's
299 close proximity to the replication origin (*oriC*, 3.9 Mb); whereas A4 exhibited a very high
300 expression level even though the integration site was the farthest from the origin among these
301 four integrants (Fig. 5B). This observation is consistent with previous studies (*32, 33*) and
302 exemplifies that position-dependent expression is affected by a myriad of known and yet-to-be-
303 elucidated factors, instead of simply the gene dosage determined by the distance from the origin.

304 With the gene transcript profiles of these isolates, we sought to elucidate the relationship
305 between gene expression and production phenotype. Some interesting observations were made.
306 For instance, predictably, isolates from Library A have lower expression of *ilvD* than isolates
307 from Libraries B and C. However, *ilvC* can be seen to be upregulated even in certain strains that
308 do not have *ilvC* insertion (A3, A4, B1). These strains have high expression levels of *alsS*, so it
309 is likely that accumulation of acetolactate, the product of AlsS, is inducing higher transcription
310 levels of *ilvC* in these strains, as has been previously observed (*49*). We also noted that there was
311 a similar trend of relative expression levels within the pathway (*ilvC* highest followed by *alsS*
312 and then *ilvD*) in the medium- and high-production isolates (A3, A4, B1, C5, C6, C7), which was
313 also seen in the plasmid strain. Most strikingly, however, there was a lack of a clear relationship
314 between gene expression levels and production. In particular, despite the fact that the top four
315 analyzed isolates (B1, C5, C6, C7) all showed a much higher expression of *ilvD* than the next
316 two, substantially lower-producing, isolates (A3, A4), it is not clear what causes C7 to perform
317 significantly better in production than the other three isolates or why differences in gene
318 expression levels across these three do not lead to differences in production. Examination of the
319 low-production isolates also provided some insights. Low expression, especially of *alsS*, appears
320 to have caused low production in A1 and A2. Intriguingly, another isolate C1 showed much
321 higher expression of the three pathway genes but did not exhibit increased production. We
322 suspect that this could be due to deleterious effects of disrupting the genes in the integration sites
323 or to poor pathway balancing. The above findings underscore the highly entangled and complex
324 nature of optimizing pathway gene expression for a specific production phenotype.

325 We have here analyzed expression only at the mRNA level and it should be noted that
326 post-transcriptional processes often have significant effects on protein abundance and activity
327 (*50*). Further study involving quantitative proteomics and enzyme activity measurements may
328 provide further insights into the mechanisms underlying the performance of the integrated strains.

329

330 *Construction and characterization of strains with all isobutanol pathway genes integrated*

331 To generate strains in which the entire isobutanol pathway is expressed from the
332 chromosome, we integrated the *adhA* and *kivD* genes into isolate C7. Because in the work of
333 Atsumi *et al.* (*38*), these genes were expressed from a high-copy-number plasmid, we expected
334 that high expression levels would be needed to achieve suitable production levels. We utilized
335 the chemically inducible chromosomal evolution (CIChE) method of Tyo *et al.* (*18*) for
336 generating multiple copies of a gene in a single integration site. In this approach, the genes of
337 interest are placed alongside an antibiotic resistance marker between 1 kb homology regions that
338 enable RecA-mediated recombination to increase the copy number in response to increases in
339 antibiotic concentration (Fig. 6A). If desired, the copy number can be stabilized by deletion of
340 the *recA* gene. We integrated the CIChE construct containing *kivD, adhA,* and the
341 chloramphenicol resistance gene *cat* into the *aslB* site, a high-expression site (*32*). Before
342 integrating this construct into C7, we also integrated it into a clean background strain, passaged
343 it into higher chloramphenicol concentrations, deleted the *recA* gene to stabilize the copy number,
344 and measured the construct copy number by qPCR of the *cat* gene (Fig. S3). These measurements
345 demonstrated that the copy number can be increased up to about 60 copies by this method. We
346 tested the production of the C7 strain with the CIChE *kivD-adhA* integration (C7 CIChE) in 50
347 μg/mL chloramphenicol and found that it produced more isobutanol than a strain with just a
348 single copy of *kivD* and *adhA* integrated into the same site. Further increasing the
349 chloramphenicol concentration did not lead to an increase in production level. The fully
350 integrated strain C7 CIChE performed very similarly to C7 pSA65, generating isobutanol titers
351 and yields lower than but close to those of the double plasmid strain (Fig. 6B,C). Noticeably, on
352 the other hand, this strain reached higher cell densities than the plasmid-bearing strains (Fig. 6D).

353 We also examined the *kivD* and *adhA* transcript levels in the fully integrated strain
354 compared to the plasmid-bearing strains. As expected, they are lower in C7 CIChE. Quite
355 surprisingly, however, we found them to be dramatically lower (more than two orders of
356 magnitude) in the integrated strain compared to those in both strains expressing *kivD*/*adhA* from
357 plasmid pSA65 (Fig. 6E). This indicates that the expression levels rendered by the high-copy
358 number plasmid are unnecessarily high. This result further demonstrates that it is feasible to
359 achieve high production through significantly lower pathway gene expression from the
360 chromosome.

361

362 **Discussion**

363

364 Synthetic biology is making strides toward solutions to some of society's most pressing
365 problems. The challenges of improving strain stability and reducing cell-to-cell variability, however,
366 still present major hurdles to overcome (*51*). Moving gene expression from plasmids to the genome
367 can help to manage these issues. Yet it remains difficult to predict the expression levels that will be
368 optimal for production and achievable by genomic integration into a particular site. Moreover, the
369 interplay between gene integration and surrounding regions on the chromosome, such as gene
370 disruption and expression level perturbation, could cause complications and additional difficulty in
371 optimizing pathway expression.

372 The new approach of screening random integration libraries developed in the present study
373 bypasses the challenges of rational selection of integration locations. The transposon integration
374 method is versatile and applicable to almost any gene of interest. Since the integration is not
375 homology-based (unlike lambda Red recombinase-based integration, for example) it can even be
376 used on genes that are already natively present in the genome at suboptimal expression levels, such
377 as *ilvC* in the work presented here. The high-throughput screening enables large libraries to be
378 assessed, making it feasible to multiplex integrations of different pathway genes.

379 The application of the approach to the isobutanol pathway demonstrates its effectiveness and
380 has generated a strain with, to our knowledge, the highest production titers yet reported in the
381 literature for an *E. coli* strain with the isobutanol pathway chromosomally integrated. This
382 chromosomally integrated strain, which achieves production performance close to that of a plasmid-
383 based strain with substantially higher pathway gene expression levels, also provides a striking
384 example of how plasmid-based expression can be unnecessarily high. We expect that the lower
385 expression levels in the integrants may be beneficial under more stressful conditions, alleviating
386 unnecessary demands on the cell. Even under the relatively optimal fermentation conditions used
387 here, the integrated strains reach higher cell densities than the plasmid-bearing strains, which may
388 in part be a result of reduced burden on these strains. Furthermore, we found this approach can take
389 advantage of beneficial gene deletions that may not be obvious choices for targeted deletion to yield
390 optimal production.

391 Our results also open intriguing questions for future investigation, the answers to which
392 would be valuable in guiding rational genome engineering endeavors. For example, it will be
393 interesting to learn what factors can lead a strain with relatively high expression levels of all pathway
394 genes, such as C1, to have such a substantially lower production level than other strains. Also
395 intriguing is the fact that, in the highest-producing isolate we identified, C7, as well as in another
396 high-producing isolate, B2, the gene integrations are located quite close to each other. It remains
397 unclear whether the good performance is due to that region having optimal expression levels, or if
398 the proximity of the constructs provides some benefit, or it has merely occurred by chance. Deep-
399 sequencing pools of droplets collected from different degrees of screening stringency could provide
400 further insights into whether proximity leads to improvements in production. Additionally, studies
401 with targeted integrations could be designed to probe potential mechanisms. Even without answers
402 to these intriguing questions regarding the mechanisms underlying the strain improvements, the
403 approach presented here, combining generation of large integration libraries by transposon

404 integration and high-throughput screening by SnoCAP, provides a streamlined and broadly
405 applicable method to generate efficient chromosomally integrated biochemical producers.
406
407
408

409
410 **Materials and Methods**
411
412 *Strains and plasmids*

413 Strains and plasmids used are listed in Table S1. JCL260, pSA65, and pSA69 (*38, 52*) were provided
414 by James Liao, UCLA. Keio strains were obtained from the *E. coli* genetic stock center (CGSC,
415 http://cgsc2.biology.yale.edu/). Plasmid pSAS31 was constructed by Scott Scholz(*33*).

416

417 *Media*

418 M9IPG, consisting of M9 salts (47.8 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.55 mM NaCl, 9.35 mM
419 NH_4Cl, 1 mM $MgSO_4$, 0.3 mM $CaCl_2$), micronutrients (2.91 nM $(NH_4)_2MoO_4$, 401.1 nM H_3BO_3,
420 30.3 nM $CoCl_2$, 9.61 nM $CuSO_4$, 51.4 nM $MnCl_2$, 6.1 nM $ZnSO_4$, 0.01 mM $FeSO_4$), thiamine HCl
421 (3.32 μM) and dextrose (D-glucose) at the stated concentrations, was used for all culturing
422 experiments. For SnoCAP screening, in both microdroplets and microplates, the glucose
423 concentration was 20 g/L, and the medium was supplemented with 3 mM L-isoleucine and 50
424 μg/mL kanamycin. For isobutanol production monocultures, the medium contained 36 g/L glucose,
425 5 g/L yeast extract, and no antibiotics were added. Precultures for both screening and production
426 cultures were carried out in LB Lennox with antibiotics appropriate to the strain. When used,
427 antibiotics were supplied at the following concentrations: ampicillin, 100 μg/mL; kanamycin, 50
428 μg/mL; tetracycline, 10 μg/mL; chloramphenicol, 50 μg/mL.

429

430 *Transposon integration library construction*

431 All primers and oligonucleotides were ordered from Integrated DNA Technologies, Inc. (IDT) and
432 are listed in Table S2. Each of the integration constructs was amplified by PCR from the pSA69
433 plasmid, adding the Tn5 mosaic ends. Primers phosph_transp_kan_for and phosph_transp_alsS_rev
434 were used to amplify *kan*-PLlacO1-*alsS* (for Libraries A, B and C).
435 phosph_transp_PLlacO1_ilvC_for and phosph_transp_ilvD_rev were used to amplify *ilvCD* and

436 introduce a PLlacO1 promoter in front of *ilvC* (for Library C). phosph_transp_PLlacO1_ilvD_for
437 and phosph_transp_ilvD_rev were used to amplify *ilvD* and introduce a PLlacO1 promoter in front
438 of *ilvD* (for Library B). The linear PCR product was then digested with both DpnI and SpeI enzymes
439 (NEB) to digest the template plasmid, phosphorylated with T4 polynucleotide kinase (NEB),
440 cleaned with a PCR clean-up kit (Qiagen) and eluted in TE buffer. The DNA was then reacted with
441 EZ-Tn5 transposase (Lucigen) according to the manufacturer's instructions. The resulting
442 transposome was then electroporated into the appropriate strain. The cells were recovered for 1.5 hr
443 with 1 mL SOC medium. Then 50 µL was used for dilution and plating on LB with kanamycin
444 plates to assess library size, and the remaining cells were grown to saturation in 100 mL selective
445 medium (either LB with 50 µg/mL kanamycin for *alsS* integration or M9IPG with 0.1 mM IPTG
446 and 3 mM lysine for *ilvCD* or *ilvD* integration). The cells were then frozen in 1 mL aliquots
447 (resuspended in fresh LB with 25% glycerol) and later thawed, washed, and grown in LB with 50
448 µg/mL kanamycin to prepare them for screening.
449
450 *SnoCAP screening*
451 SnoCAP screening for 2-KIV production, in droplet and microplate formats, was employed as
452 previously described(*36*). Stationary phase cultures in LB were used as the inocula for both formats.
453 For the microplate format, K12 Δ*ilvD* was used as the sensor strain. M9IPG with 20 g/L glucose, 3
454 mM isoleucine, and 50 µg/mL kanamycin was used as the medium for all screening co-cultures
455 except for those with the *ilvD* and *ilvCD* libraries (without *alsS* integration), in which the kanamycin
456 was omitted. For the droplet format, K12 Δ*ilvD* pSAS31, which expresses mNeonGreen, was used
457 as the sensor strain. For the droplet sorting assay, the droplet collection device was soaked in a
458 mixture of HFE-7500 oil and water for several days before use in order to improve droplet stability
459 after collection.
460

461 *Isobutanol fermentations and analysis*

462 Isobutanol cultures were grown similarly to the method previously described(*36*). Overnight
463 cultures in LB with antibiotics were diluted 1:100 v/v into 10 mL of M9IPG with 36 g/L glucose, 5
464 g/L yeast extract, and no antibiotics, in a 125-mL baffled, unvented polypropylene flasks. Cells were
465 grown for 2.5 to 3 h at 37 °C, 250 rpm, followed by induction of the pathway genes with 0.1 mM
466 IPTG. Flasks were then sealed with parafilm and incubated at 30 ^{o}C, 250 rpm. Samples were taken
467 daily to measure cell density and glucose and isobutanol concentrations. Cell density was assessed
468 by measuring the OD_{600} of 200 μL of culture, diluted into the linear range, in a VersaMax microplate
469 reader (Molecular Devices). Glucose and isobutanol were measured by high-performance liquid
470 chromatography (HPLC), as previously described[37]. In the case of cultures in which RNA was
471 harvested, 0.5 mL of cells were harvested with RNAprotect Bacteria Reagent (Qiagen) 3-4 h after
472 induction. RNA was isolated using an RNA mini kit (Qiagen) according to the manufacturer's
473 instructions, including on-column DNAse I digestion. An additional digestion using Turbo DNA-
474 free Kit (Invitrogen) was performed after purification to eliminate any remaining genomic DNA.
475 The RNA was then reverse transcribed to cDNA using MultiScribe Reverse Transcriptase
476 (Invitrogen) and approximately 400 ng RNA per 20 μL reaction, with the following temperature
477 profile: 25 ^{o}C, 10 min; 37 ^{o}C, 120 min, 85 ^{o}C, 5 min. A no reverse transcriptase control was carried
478 out for each sample and checked with at least one set of primers during the real-time PCR analysis
479 to ensure no significant residual genomic DNA. qPCR was conducted on the cDNA using SYBR
480 Green qPCR Master Mix (Life Technologies) with 20 μL reactions. Reactions were run in 96-well
481 plates on a 7900HT Fast Real-Time PCR Machine (Applied Biosystems), courtesy of the University
482 of Michigan Advanced Genomics Core. The PCR program was as follows: 50 ^{o}C, 2 min; 95 ^{o}C, 10
483 min; 40 cycles of 95 ^{o}C for 15 s, 60 ^{o}C for 1 min. Primers used are listed in Table S2. Data were
484 analyzed using the $2^{-\Delta\Delta CT}$ method (*53*).

485
486

487 *kivD and adhA integration*
488
489 A CIChE construct consisting of *kivD, adhA*, and a chloramphenicol resistance gene, *cat*, flanked
490 by 1 kb homology regions, was previously generated and integrated into the *aslB* locus of the
491 NV3r1 strain by λ-Red recombineering (*36*). Following the method of Tyo *et al.* (*18*), we
492 passaged this strain into increasing concentrations of chloramphenicol and then deleted the *recA*
493 gene by P1 transduction from donor strain BW 26,547 Δ*recA*::*kan* Lambda *recA*+ (obtained from
494 the Coli Genetic Stock Center (CGSC)). Transductants were selected on LB plates with 50 μg/mL
495 kanamycin and the corresponding concentration of chloramphenicol. Genomic DNA was isolated
496 using DNEasy Blood & Tissue Kit (Qiagen), and the copy number was determined by qPCR using
497 the same primers as Tyo *et al.* (*18*), with one set for the *cat* gene and one for the *bioA* gene, which
498 is used as a single copy reference gene. Absolute copy number was determined by comparison to a
499 strain with a single copy of *cat* integrated into the *intC* locus (NV3r1 Δ*intC::yfp-cat*). To transfer
500 this integration into the C7 strain generated in this study, we prepared P1 lysates from NV3r1
501 (*recA*+) with the *kivD, adhA* CIChE construct after growing on 80 μg/mL chloramphenicol and
502 transduced it into the C7 strain, selecting on LB plates with 40 μg/mL chloramphenicol. Resulting
503 colonies were isolation streaked twice on LB agar with 40 μg/mL chloramphenicol and 0.8 mM
504 sodium citrate.
505
506
507

508
509 **References**
510

511 1. B. M. Woolston, S. Edgar, G. Stephanopoulos, Metabolic Engineering: Past and Future.
512 *Annual Review of Chemical and Biomolecular Engineering* **4**, 259–288 (2013).
513 2. J. Nielsen, Jay D. Keasling, Engineering cellular metabolism. *Cell* **164**, 1185-1197 (2016).
514 3. M. Popov, S. Petrov, G. Nacheva, I. Ivanov, U. Reichl, Effects of a recombinant gene
515 expression on ColE1-like plasmid segregation in *Escherichia coli*. *BMC Biotechnology* **11**,
516 18 (2011).
517 4. W. E. Bentley, D. S. Kompala, Plasmid instability in batch cultures of recombinant
518 bacteria. A laboratory experiment. *Chemical Engineering Education* **24**, 168-172 (1990).
519 5. J. H. Seo, J. E. Bailey, A segregated model for plasmid content and product synthesis in
520 unstable binary fission recombinant organisms. *Biotechnology and Bioengineering* **27**,
521 156-166 (1985).
522 6. R. Siegel, D. D. Ryu, Kinetic study of instability of recombinant plasmid pPLc23trpAl in
523 *E. coli* using two-stage continuous culture system. *Biotechnology and Bioengineering* **27**,
524 28-33 (1985).
525 7. T. Imanaka, H. Tsunekawa, S. Aiba, Phenotypic stability of trp operon recombinant
526 plasmids in *Escherichia coli*. *Microbiology* **118**, 253-261 (1980).
527 8. T. Imanaka, S. Aiba, A perspective on the application of genetic engineering: stability of
528 recombinant plasmid. *Annals of the New York Academy of Sciences* **369**, 1-14 (1981).
529 9. D. Ollis, Industrial fermentations with (unstable) recombinant cultures. *Philosophical*
530 *Transactions of the Royal Society of London. B, Biological Sciences* **297**, 617-629 (1982).
531 10. K. P. Friehs, in *"Plasmid copy number and plasmid stability"* in *New Trends and*
532 *Developments in Biochemical Engineering*. (Springer, 2004), pp. 47-82.
533 11. B. D. Ensley, Stability of Recombinant Plasmids in Industrial Microorganisms. *Critical*
534 *Reviews in Biotechnology* **4**, 263-277 (1986).
535 12. D. K. Summers, D. J. Sherratt, Multimerization of high copy number plasmids causes
536 instability: ColE 1 encodes a determinant essential for plasmid monomerization and
537 stability. *Cell* **36**, 1097-1103 (1984).
538 13. C. Field, D. Summers, Multicopy plasmid stability: revisiting the dimer catastrophe.
539 *Journal of Theoretical Biology* **291**, 119-127 (2011).
540 14. M. Jahn, C. Vorpahl, T. Hübschmann, H. Harms, S. Müller, Copy number variability of
541 expression plasmids determined by cell sorting and Droplet Digital PCR. *Microbial Cell*
542 *Factories* **15**, 211 (2016).
543 15. K. M. Münch *et al.*, Polar fixation of plasmids during recombinant protein production in
544 Bacillus megaterium results in population heterogeneity. *Applied and Environmental*
545 *Microbiology* **81**, 5976-5986 (2015).
546 16. F. W. Studier, Protein production by auto-induction in high-density shaking cultures.
547 *Protein Expression and Purification* **41**, 207-234 (2005).
548 17. C. Mignon, R. Sodoyer, B. Werle, Antibiotic-free selection in biotherapeutics: now and
549 forever. *Pathogens* **4**, 157-181 (2015).
550 18. K. E. J. Tyo, P. K. Ajikumar, G. Stephanopoulos, Stabilized gene duplication enables long-
551 term selection-free heterologous pathway expression. *Nature Biotechnology* **27**, 760-765
552 (2009).
553 19. G. Striedner *et al.*, Plasmid-free T7-based *Escherichia coli* expression systems.
554 *Biotechnology and Bioengineering* **105**, 786-794 (2010).
555 20. J. Mairhofer, T. Scharl, K. Marisch, M. Cserjan-Puschmann, G. Striedner, Comparative
556 transcription profiling and in-depth characterization of plasmid-based and plasmid-free

557 *Escherichia coli* expression systems under production conditions. *Applied and*
558 *Environmental Microbiology* **79**, 3802-3812 (2013).
559 21. M. Madyagol *et al.*, Gene replacement techniques for *Escherichia coli* genome
560 modification. *Folia Microbiologica* **56**, 253-263 (2011).
561 22. B. Ou, C. Garcia, Y. Wang, W. Zhang, G. Zhu, Techniques for chromosomal integration
562 and expression optimization in *Escherichia coli*. *Biotechnology and Bioengineering* **115**,
563 (2018).
564 23. L. Li, X. Liu, K. Wei, Y. Lu, W. Jiang, Synthetic biology approaches for chromosomal
565 integration of genes and pathways in industrial microbial systems. *Biotechnology Advances*
566 **37**, 730-745 (2019).
567 24. Y. Zhang, F. Buchholz, J. P. Muyrers, A. F. Stewart, A new logic for DNA engineering
568 using recombination in *Escherichia coli*. *Nature Genetics* **20**, 123-128 (1998).
569 25. D. Yu, H. M. Ellis, E.-C. Lee, N. A. Jenkins, N. G. Copeland, An efficient recombination
570 system for chromosome engineering in *Escherichia coli*. *Proceedings of the National*
571 *Academy of Sciences* **97**, 5978-5983 (2000).
572 26. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in
573 *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of*
574 *Sciences* **97**, 6640-6645 (2000).
575 27. N. D. Grindley, K. L. Whiteson, P. A. Rice, Mechanisms of site-specific recombination.
576 *Annual Review of Biochemistry* **75**, 567-605 (2006).
577 28. N. I. Minaeva *et al.*, Dual-In/Out strategy for genes integration into bacterial chromosome:
578 a novel approach to step-by-step construction of plasmid-less marker-less recombinant *E.*
579 *coli* strains with predesigned genome structure. *BMC Biotechnology* **8**, 63 (2008).
580 29. J. Yin *et al.*, Effects of chromosomal gene copy number and locations on
581 polyhydroxyalkanoate synthesis by *Escherichia coli* and *Halomonas* sp. *Applied*
582 *Microbiology and Biotechnology* **99**, 5523-5534 (2015).
583 30. J. R. Beckwith, E. R. Signer, W. Epstein, in *Cold Spring Harbor Symposia on Quantitative*
584 *Biology*. (Cold Spring Harbor Laboratory Press, 1966), vol. 31, pp. 393-401.
585 31. M. B. Schmid, J. R. Roth, Gene location affects expression level in *Salmonella*
586 *typhimurium*. *Journal of Bacteriology* **169**, 2872-2875 (1987).
587 32. J. A. Bryant, L. E. Sellars, S. J. Busby, D. J. Lee, Chromosome position effects on gene
588 expression in *Escherichia coli* K-12. *Nucleic Acids Research* **42**, 11383-11392 (2014).
589 33. S. A. Scholz *et al.*, High-resolution mapping of the *Escherichia coli* chromosome reveals
590 positions of high and low transcription. *Cell Systems* **8**, 212-225 (2019).
591 34. C. Sousa, V. de Lorenzo, A. Cebolla, Modulation of gene expression through
592 chromosomal positioning in *Escherichia coli*. *Microbiology* **143**, 2071-2078 (1997).
593 35. A. Loeschcke *et al.*, TREX: a universal tool for the transfer and expression of biosynthetic
594 pathways in bacteria. *ACS Synthetic Biology* **2**, 22-33 (2012).
595 36. T. E. Saleski *et al.*, Syntrophic co-culture amplification of production phenotype for high-
596 throughput screening of microbial strain libraries. *Metabolic Engineering* **54**, 232-243
597 (2019).
598 37. C.-T. Chen, J. C. Liao, Frontiers in microbial 1-butanol and isobutanol production. *FEMS*
599 *Microbiology Letters* **363**, fnw020 (2016).
600 38. S. Atsumi, T. Hanai, J. C. Liao, Non-fermentative pathways for synthesis of branched-
601 chain higher alcohols as biofuels. *Nature* **451**, 86 (2008).
602 39. A. Baez, K.-M. Cho, J. C. Liao, High-flux isobutanol production using engineered
603 *Escherichia coli*: a bioreactor study with in situ product removal. *Applied Microbiology*
604 *and Biotechnology* **90**, 1681-1690 (2011).
605 40. S. Bastian, Liu, X., Meyerowitz, J.T., Snow, C.D., Chen, M.M. and Arnold, F.H.,
606 Engineered ketol-acid reductoisomerase and alcohol dehydrogenase enable anaerobic 2-

607 methylpropan-1-ol production at theoretical yield in Escherichia coli. *Metabolic*
608 *Engineering* **13**, 345-352 (2011).
609 41. I. N. Ghosh, Martien, J., Hebert, A.S., Zhang, Y., Coon, J.J., Amador-Noguez, D. and
610 Landick, R., OptSSeq explores enzyme expression and function landscapes to maximize
611 isobutanol production rate. *Metabolic Engineering* **52**, 324-340 (2019).
612 42. H. Akita, N. Nakashima, T. Hoshino, Bacterial production of isobutanol without expensive
613 reagents. *Applied Microbiology and Biotechnology* **99**, 991-999 (2015).
614 43. M. C. Bassalo *et al.*, Rapid and efficient one-step metabolic pathway integration in *E. coli*.
615 *ACS Synthetic Biology* **5**, 561-568 (2016).
616 44. J.-C. Baret *et al.*, Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell
617 sorting based on enzymatic activity. *Lab on a Chip* **9**, 1850-1858 (2009).
618 45. R. C. Clowes, Molecular structure of bacterial plasmids. *Bacteriological Reviews* **36**, 361
619 (1972).
620 46. G. Ebersbach, K. Gerdes, Plasmid segregation mechanisms. *Annual Review of Genetics* **39**,
621 453-479 (2005).
622 47. A. H. Kachroo, M. Jayaram, P. A. Rowley, Metabolic engineering without plasmids.
623 *Nature Biotechnology* **27**, 729 (2009).
624 48. J. D. Keasling, Gene-expression tools for the metabolic engineering of bacteria. *Trends in*
625 *Biotechnology* **17**, 452-460 (1999).
626 49. S. M. Arfin, B. Ratzkin, H. Umbarger, The metabolism of valine and isoleucine in
627 *Escherichia coli.* XVII. The role of induction in the depression of acetohydroxy acid
628 isomeroreductase. *Biochemical and Biophysical Research Communications* **37**, 902-908
629 (1969).
630 50. T. Maier, M. Güell, L. Serrano, Correlation of mRNA and protein in complex biological
631 samples. *FEBS Letters* **583**, 3966-3973 (2009).
632 51. T. Wang, M. J. Dunlop, Controlling and exploiting cell-to-cell variation in metabolic
633 engineering. *Current Opinion in Biotechnology* **57**, 10-16 (2019).
634 52. S. Atsumi *et al.*, Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by
635 comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Applied*
636 *Microbiology and Biotechnology* **85**, 651-657 (2010).
637 53. K. J. Livak, T. D. Schmittgen, Analysis of relative gene expression data using real-time
638 quantitative PCR and the 2(T)(-Delta Delta C) method. *Methods* **25**, 402-408 (2001).
639

640 **Supplemental References**
641
642 1. Atsumi, S., Hanai, T. & Liao, J.C. Non-fermentative pathways for synthesis of branched-
643 chain higher alcohols as biofuels. *Nature* **451**, 86 (2008).
644 2. Saleski, T.E. et al. Syntrophic co-culture amplification of production phenotype for high-
645 throughput screening of microbial strain libraries. *Metabolic Engineering* **54**, 232-243
646 (2019).
647 3. Smith, K.M. & Liao, J.C. An evolutionary strategy for isobutanol production strain
648 development in *Escherichia coli*. *Metabolic Engineering* **13**, 674-681 (2011).
649 4. Elowitz, M.B., Levine, A.J., Siggia, E.D. & Swain, P.S. Stochastic gene expression in a
650 single cell. *Science* **297**, 1183-1186 (2002).
651 5. Atsumi, S. et al. Engineering the isobutanol biosynthetic pathway in *Escherichia coli* by
652 comparison of three aldehyde reductase/alcohol dehydrogenase genes. *Applied*
653 *Microbiology and Biotechnology* **85**, 651-657 (2010).
654 6. Scholz, S.A. et al. High-resolution mapping of the *Escherichia coli* chromosome reveals
655 positions of high and low transcription. *Cell Systems* **8**, 212-225 (2019).
656 7. Zhou, K. et al. Novel reference genes for quantifying transcriptional responses of
657 *Escherichia coli* to protein overexpression by quantitative PCR. *BMC Molecular Biology*
658 **12**, 18-18 (2011).
659
660
661
662
663
664

Acknowledgments

General: The authors thank Scott Scholz and Peter Freddolino for valuable discussions regarding position-dependent gene expression variation and transposon footprinting, as well as Mark Burns, Lola Eniola-Adefeso, and Chris Barr for the use of equipment in their laboratories. The isobutanol strains and plasmids were originally provided by James Liao (UCLA, formerly). The CIChE construct was constructed using the pTGD plasmid provided by Keith Tyo (Northwestern U).

Funding: This work was supported by the USDA AFRI NIFA Fellowships Grant Program (Grant no. 2016-67011-24725).

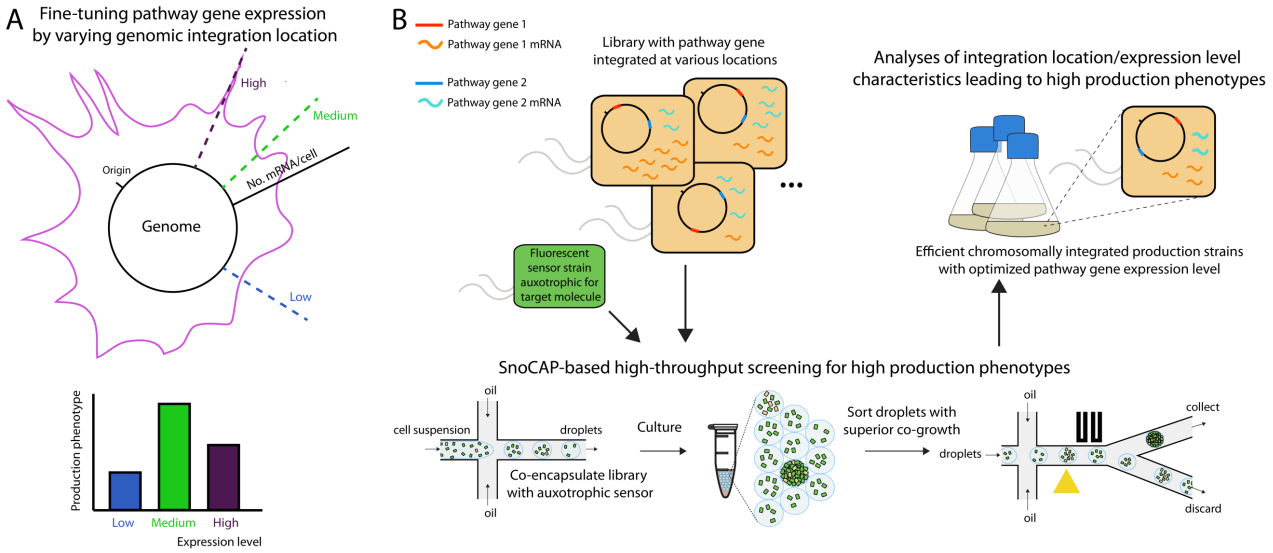
Author contributions: T.E.S. and X.N.L. conceived and designed the study. T.E.S. and A.K. constructed the integration libraries. M.T.C. and K.K. developed the droplet sorting platform. T.E.S. and M.T.C. performed the library screening. T.E.S. and D.N.C. characterized the library isolates and fully integrated strains. T.E.S. and X.N.L. wrote the manuscript. All authors discussed the results and commented on the manuscript.

Data Availability: All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Competing interests: The authors declare that they have no competing interests.

692 **Figures and Tables**

693



694
695

696 **Figure 1. New approach for optimizing chromosomal heterologous gene expression by**
697 **position-dependent expression variation.** (A) Gene expression level varies depending on
698 distance from the origin of replication, as well as other position-dependent factors. The
699 expression level of a heterologous pathway gene affects the production phenotype of a strain.
700 Increasing the expression level of a gene can improve production level by increasing flux through
701 the pathway, but avoiding too high of an expression level can also provide benefits such as
702 reducing the buildup of toxic intermediates and avoiding unnecessary metabolic burden. (B)
703 Libraries with pathway genes integrated into various genome locations can be screened for
704 production phenotype using syntrophic co-culture amplification of production (SnoCAP) (*36*). A
705 cross-feeding co-culture is configured consisting of a secretor strain producing a target molecule
706 and a sensor strain that cannot synthesize the target molecule autonomously. The co-cultures are
707 compartmentalized in microfluidic droplets, incubated to allow co-growth, and then sorted to
708 identify the strains with high production phenotypes.

709

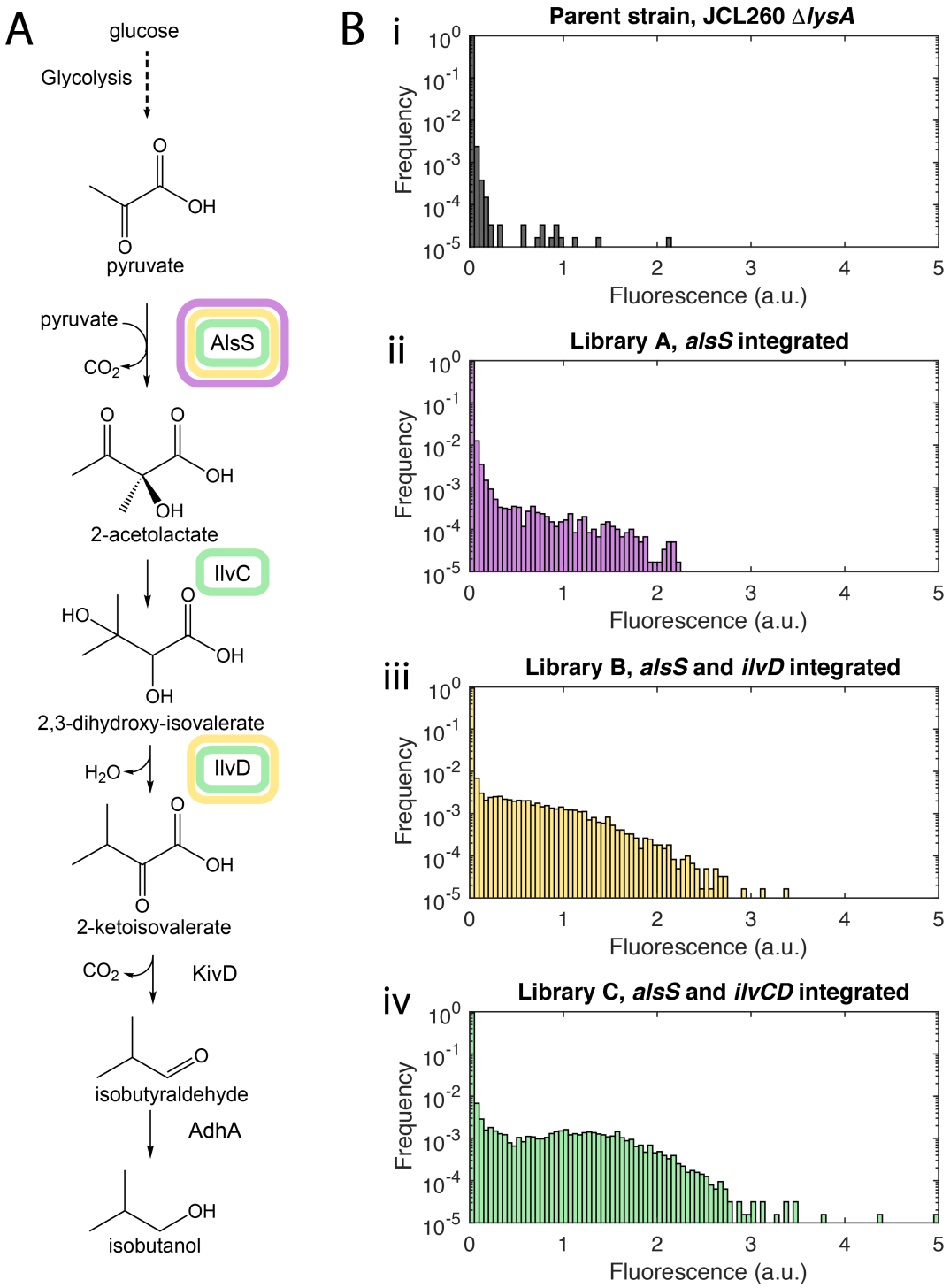


Figure 2. Screening of libraries with *alsS*, *ilvCD* integrated by Tn5. (A) Isobutanol production pathway (*36, 52*). Acetolactate synthase (AlsS), ketol-acid reductoisomerase (IlvC), dihydroxy-acid dehydratase (IlvD), 2-ketoacid decarboxylase (KivD), and alcohol dehydrogenase (AdhA). Colored boxes correspond to the integrations in Libraries A (magenta), B (yellow), and C (green). (B) Histograms comparing fluorescence signal from droplets containing co-cultures of fluorescent sensor strain with unlabeled secretor strain libraries, compared to the parent strain (i). The libraries were constructed by Tn5 integration of *kan*-PLlacO1-*alsS* (Library A) (ii), PLlacO1-*ilvD* followed by *kan*-PLlacO1-*alsS* (Library B) (iii), or PLlacO1-*ilvCD* followed by *kan*-PLlacO1-*alsS* (Library C) (iv). Each histogram comprises data from the analysis of about 60,000 droplets.

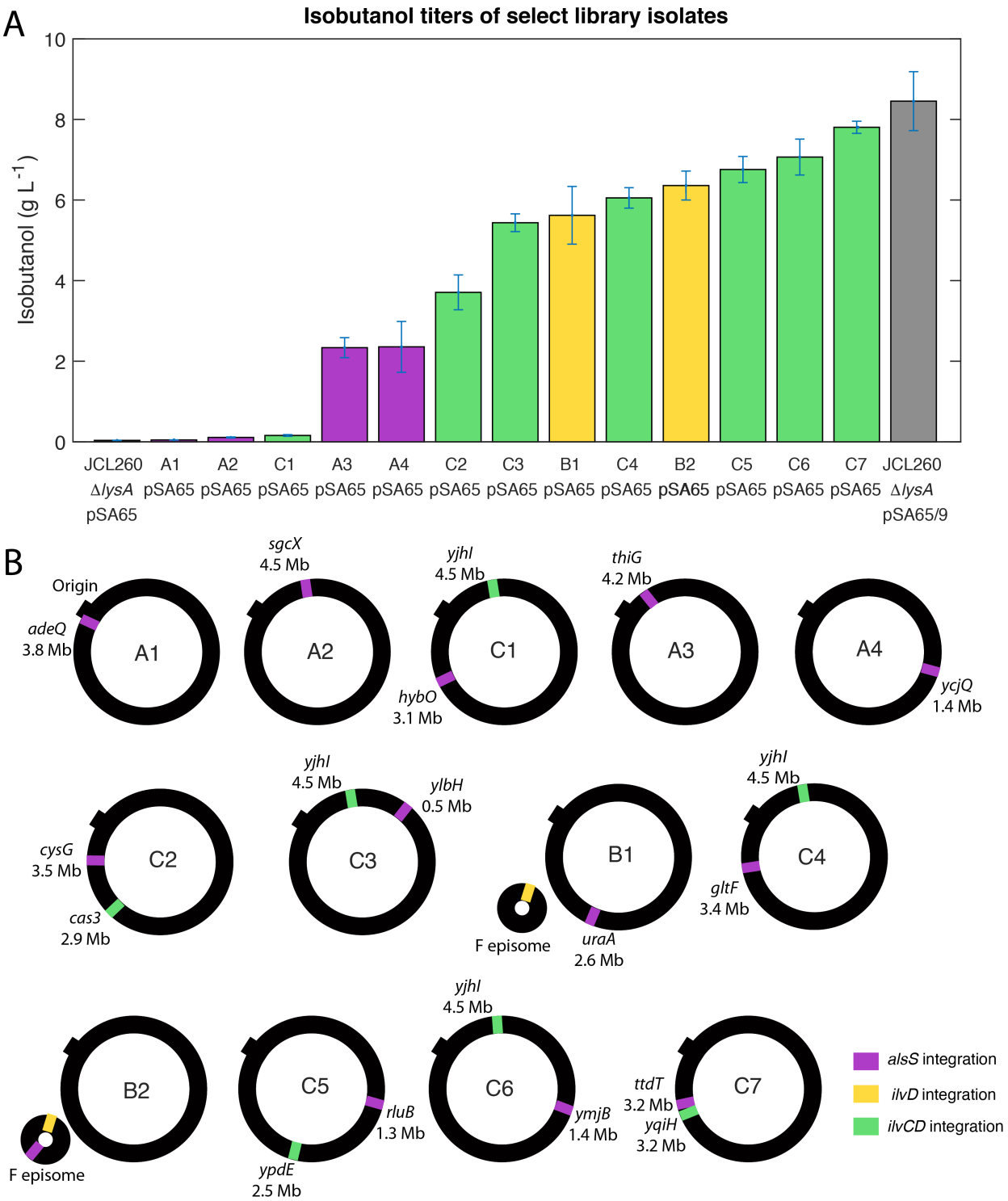


Figure 3. Production phenotype and gene insertion locations of select library isolates. (A) Isobutanol titers of library isolates displaying various production levels. Samples were taken 48 h after induction of the pathway genes. Each of the isolates has been transformed with the pSA65 plasmid expressing *kivD* and *adhA*. Parent strain JCL260 Δ*lysA* and JCL260 Δ*lysA* pSA69, which expresses the *alsS* and *ilvCD* genes from the pSA69 plasmid, each also carrying pSA65, are included for comparison. Error bars represent the standard deviation of three biological replicates. (B) Gene integration locations of the isolates whose production is shown in (A).

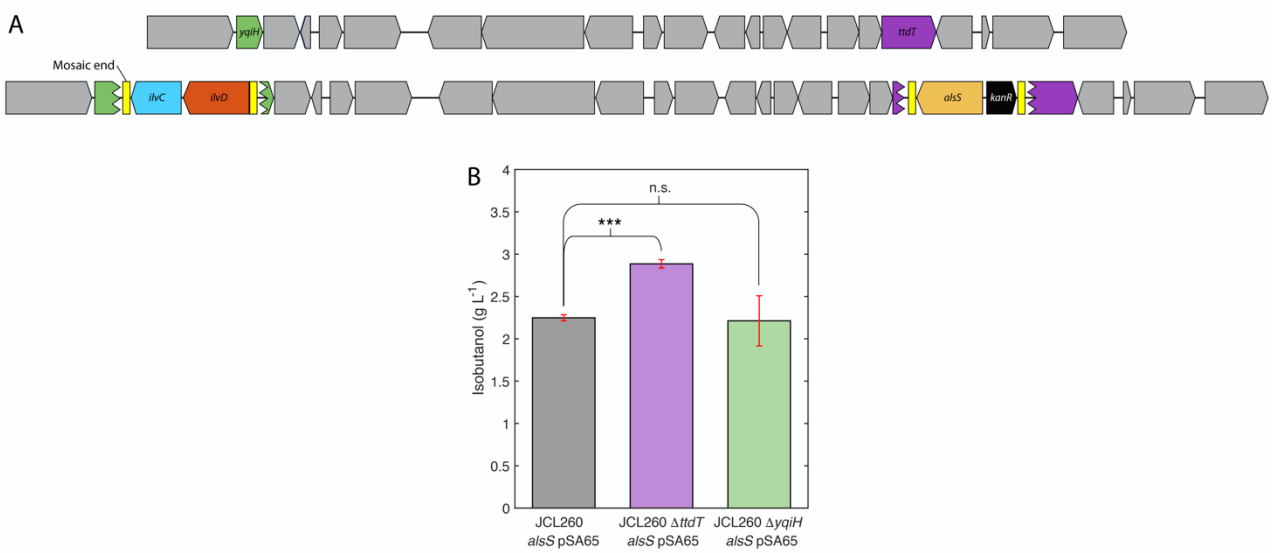


Figure 4. Examination of the effect of gene disruptions incurred during the Tn5 integrations in the C7 isolate on production phenotype. (A) Schematic of the locations of *kan*-PLlacO1-*alsS* and PLlacO1-*ilvCD* integrations in isolate C7, determined by transposon footprinting. (B) Isobutanol production of strains with *alsS* genomically integrated, containing plasmid pSA65 encoding *kivD* and *adhA*, and possessing deletions of either *ttdT* or *yqiH*, compared to the same strain without either deletion. Data shown are from samples taken 72 h after induction with IPTG. Error bars in (B) represent the standard deviation of three biological replicates. ***: significant difference at *P*-value < 0.0001 using an unpaired t-test. n.s.: not significant.

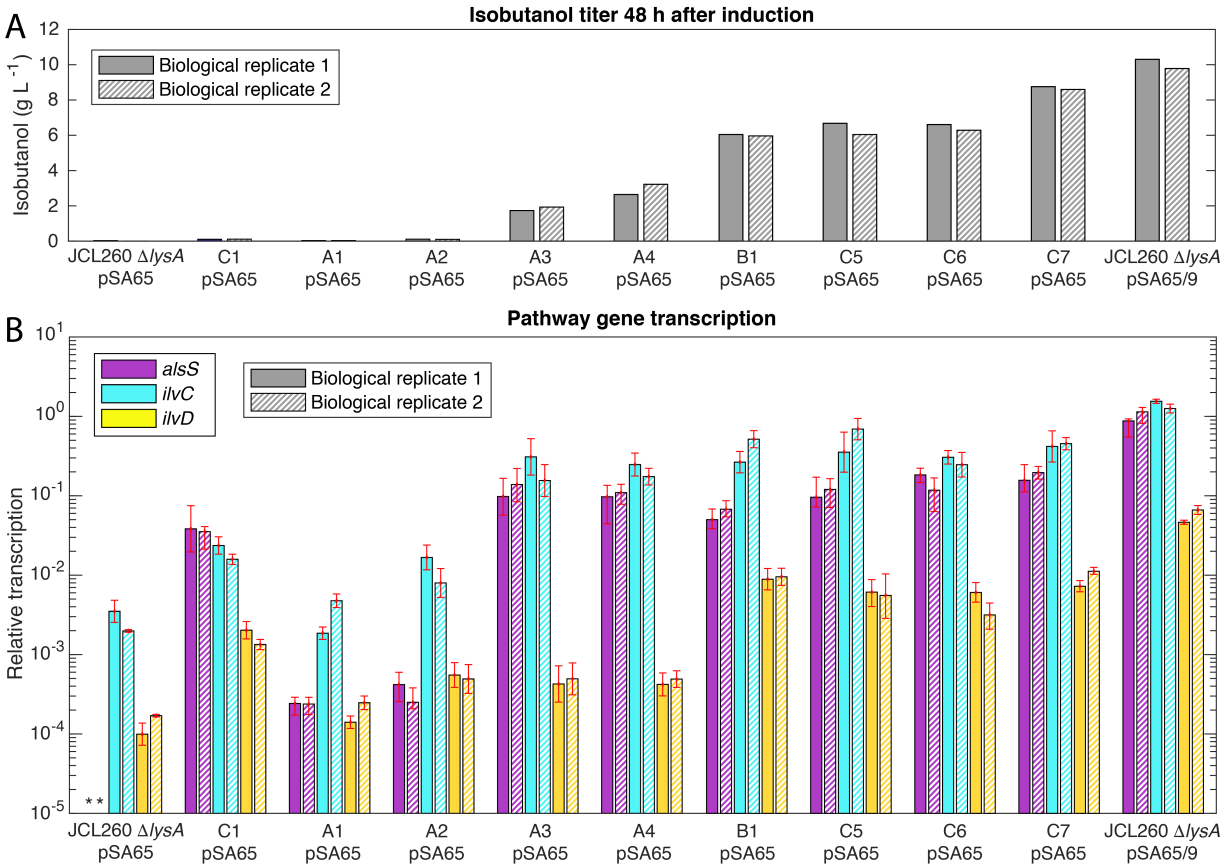


Figure 5. Investigation of the expression levels underlying production phenotypes of select library isolates. (A) Isobutanol production 48 h after induction of the cultures from which the RNA had been harvested. (B) Expression levels of *alsS*, *ilvC*, and *ilvD*, relative to *alsS* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 3 h after induction with IPTG. * indicates no amplification during the qPCR reaction. Data from two biological replicate cultures of each strain are shown. Error bars in (B) represent the standard deviation of three technical replicates.

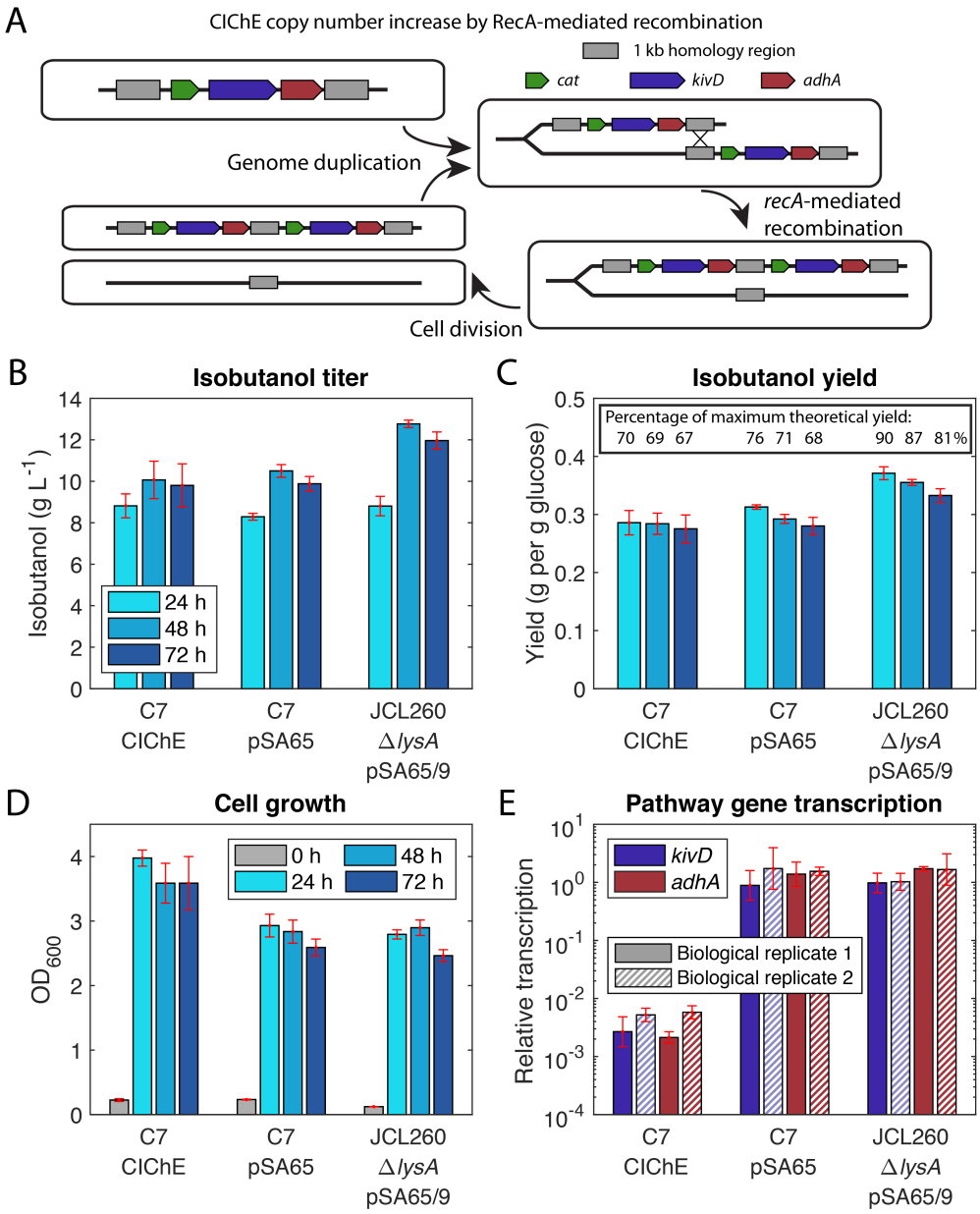


Figure 6. Fully integrated strain performance compared to strains with plasmids. (A) Overview of the chemically inducible chromosomal evolution (CIChE) approach for gene copy number amplification. A construct containing the genes of interest, *kivD* and *adhA*, along with the chloramphenicol resistance marker, *cat*, flanked by two identical 1 kb homology regions, is integrated into the *E. coli* chromosome. RecA-mediated recombination of the homology regions results in some daughter cells having higher copy number of the integrated genes and thus able to tolerate higher chloramphenicol concentrations. (B) Isobutanol production by strain C7 after integration of *kivD-adhA* by CIChE, compared to C7 pSA65, which has *alsS* and *ilvCD* integrated into the genome and *kivD* and *adhA* expressed from the pSA65 plasmid, and JCL260 Δ*lysA* pSA65/9 which has all five pathway genes expressed from the plasmids. Error bars represent the standard deviation of three biological replicates. (C) The yields of the cultures based on glucose consumed. The inset in (C) lists the percentage of the theoretical maximum yield (0.41 g isobutanol per gram glucose). (D) The cell densities of the production cultures whose performance is displayed in (B) and (C). (E) Expression levels of *kivD* and *adhA*, relative to *kivD* level in JCL260 Δ*lysA* pSA65/9. RNA was harvested 4 h after induction with IPTG. Error bars represent the standard deviation of three technical replicates.

771 **Table 1.** Gene integration locations.
772

Strain	Integration construct	Gene	Coordinate in gene of integration location	Direction relative to gene integrated into
A1	*kan-alsS*	*adeQ*	436	Opposite
A2	*kan-alsS*	*sgcX*	111	Opposite
C1	*kan-alsS*	*hybO*	812	Opposite
C1	*ilvCD*	*yjhI*	231	Same
A3	*kan-alsS*	*thiG*	73	Same
A4	*kan-alsS*	*ycjQ*	364	Opposite
C2	*kan-alsS*	*cysG*	563	Opposite
C2	*ilvCD*	*cas3*	1,387	Same
C3	*kan-alsS*	*ylbH*	205	Opposite
C3	*ilvCD*	*yjhI*	231	Same
B1	*kan-alsS*	*uraA*	659	Same
B1	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C4	*kan-alsS*	*gltF*	313	Opposite
C4	*ilvCD*	*yjhI*	231	Same
B2	*kan-alsS*	Upstream of *psiB* on F plasmid	218 bp upstream	Opposite of *psiB*
B2	*ilvD*	gamma delta (Tn1000) region of F plasmid	426 bp into *tnpX*	Same as *tnpX*
C5	*kan-alsS*	*rluB*	794	Same
C5	*ilvCD*	*ypdE*	76	Same
C6	*kan-alsS*	*ymjB*	7	Opposite
C6	*ilvCD*	*yjhI*	231	Same
C7	*kan-alsS*	*ttdT*	114	Opposite
C7	*ilvCD*	*yqiH*	512	Opposite

773
774